Parameter estimation using data assimilation in an atmospheric general circulation model: From a perfect towards the real world

Sebastian Schirber[1], Daniel Klocke[1,2], Robert Pincus[3], Johannes Quaas[4] and Jeffrey L. Anderson[5]

Jeffrey L. Anderson, National Center for Atmospheric Research, 1850 Table Mesa Drive, Boulder, CO 80307-3000, USA. (jla@ucar.edu)

Daniel Klocke, ECMWF, Shinfield Park, Reading, RG2 9AX, United Kingdom. (daniel.klocke@ecmwf.int)

Robert Pincus, CIRES, University of Colorado, 216 UCB, Boulder, CO 80307-3000, USA. (robert.pincus@colorado.edu)

Johannes Quaas, Institut für Meteorologie, Universität Leipzig, Stephanstr. 3, 04103 Leipzig, Germany. (johannes.quaas@uni-leipzig.de)

Sebastian Schirber, Max-Planck-Institut für Meteorologie, Bundesstrasse 53, 20146 Hamburg, Germany. (sebastian.schirber@zmaw.de)

[1]Max-Planck-Institut für Meteorologie,

Abstract. This study explores the viability of parameter estimation in the comprehensive general circulation model ECHAM6 using ensemble Kalman filter data assimilation techniques. Four closure parameters of the cumulus-convection scheme are estimated using increasingly less-idealized scenarios ranging from perfect-model experiments to the assimilation of conventional observations. Updated parameter values from experiments with real observations are used to assess the error of the model state on short, 6 hour forecasts, and on climatological timescales.

Hamburg, Germany.

[2]European Centre for Medium Range Weather Forecasts, Reading, United Kingdom.

[3]Cooperative Institute for Research in Environmental Sciences, University of Colorado and NOAA Earth System Research Laboratory, Physical Sciences Division, Boulder, CO, USA.

[4]Universität Leipzig, Leipzig, Germany.

[5]National Center for Atmospheric Research, Boulder, CO, USA.

All parameters converge to their default values in single parameter perfect-model experiments. Estimating parameters simultaneously has a neutral effect but applying an imperfect model deteriorates the assimilation performance. With real observations, single parameter estimation generates the default parameter value in one case, convergence to different parameter values in two cases and diverges in the fourth case. The implementation of the two converging parameters influences the model state: While the estimated parameter values lead to an overall error reduction on short timescales, the error of the model state increases on climatological timescales.

1. Estimating parameters for fast processes in a climate model

Current state-of-the-art climate models are truncated at fairly coarse spatial resolutions, typically of the order of 100 km. Many atmospheric processes with significant impacts on the large-scale state, including precipitation formation, radiative transfer, turbulence, and convection occur at much smaller scales. In truncated models these processes must be represented by so-called "parameterizations", statistical formulations that determine the impact of these processes on the large-scale state in terms of the state itself.

Parameterizations require closure parameters. The optimal values of these parameters, which may depend on the model's spatial and temporal resolution [Tiedtke, 1989], are determined during model "tuning", usually by adjusting parameters so that the mean model state matches climatological observations as closely as possible [Randall and Wielicki, 1997; Mauritsen et al., 2012]. The choice of parameter values plays an important role in climate prediction scenarios since it is parameters, rather than initial conditions, which determine the model's climate [Murphy et al., 2004]. Model tuning is thus a necessary but subjective and arbitrary process in the development of a climate model.

Model tuning, as normally performed, has several disadvantages. The iterative process of modifying a parameter value, running a climate simulation, comparing the model output to observations, and readjusting the parameter value is both computationally expensive and labour-intensive. The time is usually spent by central members of the model-development team since finding an optimal parameter set requires deep knowledge of the model and its parameterizations. The process is also somewhat arbitrary. Tuning is normally guided by a subjectively chosen set of parameters and targets, i.e. features of

the climate system, on which the model is calibrated. But tuning need not lead to unique parameter choices if the target can be reached by adjusting more than one parameter, i.e. if several cloud closure parameters impact the radiation budget. Moreover, the best climate state may well be achieved by compensating errors in different processes rather than by best simulating a certain physical process.

A variety of more systematic approaches to tuning has been explored. One brute-force possibility is to systematically explore the parameter space in sensitivity experiments [Allen, 1999; Knutti et al., 2002; Murphy et al., 2004; Klocke et al., 2011] although this is computationally expensive even for a modest number of free parameters. The parameter space can be explored more selectively using for example Monte Carlo Markov chains [MCMC; Jackson et al., 2008; Järvinen et al., 2010]. Both of these techniques, like traditional tuning methods, use metrics related to climatological agreement with observations. But many of the parameters normally adjusted during tuning are related to fast processes such as convection and radiation. This suggests that model sensitivity to these parameters should be evident even in very short integrations such as those used in numerical weather prediction (NWP) [Rodwell and Palmer, 2007].

NWP relies on data assimilation, an optimal blending of prior information (usually short-term forecasts) with observations, to produce optimal initial conditions for subsequent forecasts. Assimilation can be naturally extended to simultaneous estimation of state and parameter values. Parameters can be estimated using "state space augmentation" [Derber, 1989; Anderson, 2001; Norris et al., 2007], i.e. by extending the state vector to include the desired parameters which are then updated along with the physical state. Ensemble methods such as the ensemble Kalman filter (EnKF) due to Evensen [2003] are

particularly alluring for parameter estimation because covariances are sampled from the ensemble and don't need be specified.

Simultaneous parameter and state estimation has yielded promising results in low order models [Anderson, 2001; Annan and Hargreaves, 2004] and in simplified primitive equation atmospheric global models [Annan et al., 2005]. The technique has also been applied to a limited-domain NWP model of operational complexity [MM5; Aksoy et al., 2006b]. Better parameter estimates can lead to better models: Hu et al. [2010] estimated two parameters of a boundary layer scheme and the updated parameter values led to reduced model errors. But the greatest successes have been in simplified settings. Aksoy et al. [2006a]; Tong and Xue [2008] show that estimating several parameters simultaneously often degrades the estimation performance of the individual parameters.

In the present study, we apply sequential data assimilation techniques to a climate model ensemble to estimate four closure parameters of the cumulus-convection scheme. We focus on cloud and convection parameters for several reasons: a) they are important for the representation of weather and climate; b) those parameters remain very uncertain and are often used to adjust a models' weather or climate to best fit observations; c) cloud processes act on timescales short enough to potentially yield a successful parameter estimation; d) the response of clouds to an external forcing is a large contributor to the uncertainty in estimates of climate sensitivity [Soden and Held, 2006; Bony et al., 2006];

Here we present a series of experiments with decreasing degrees of idealization, from experiments with synthetic observations on a homogeneous observation network assuming a perfect-forecast model, to real observations and a consequently imperfect model (Table 1). This hierarchical approach demonstrates first that the observations we assimilate

inform the parameters we try to estimate on short timescales. By moving incrementally towards the real world the possibilities of parameter estimations in GCMs are highlighted, but also instructive limitations are demonstrated when those concepts are transferred to imperfect models and incomplete observations.

Section 2 introduces the methodology and the model employed in this study. In the following section, perfect-model experiments with single parameter estimation are used to develop a potentially successful data assimilation setup and to demonstrate the validity of the approach. We add complexity incrementally, first by simultaneously estimating multiple parameters and then by introducing imperfection to the forecast model in section 4. Experiments with real observations are described in section 5 and the performance of the updated parameter values is assessed in short forecasts and climatological model runs in section 6 before we end with a summary of the results and a conclusion.

2. A climate model making short forecasts

We use the climate model ECHAM6 [Stevens et al., 2012] with a horizontal triangular truncation T31 and 19 vertical levels on prescribed sea surface temperatures (SST) and sea ice concentration (SIC). The model is run in NWP mode, by initializing short forecasts with analysis of the atmospheric state. The initial conditions for each forecast are created with the Data Assimilation Research Testbed [DART; Anderson et al., 2009], developed at the National Center for Atmospheric Research (NCAR). We apply the ensemble adjustment Kalman filter [EAKF; Anderson, 2001] to a 90 member ensemble.

The parameters to be estimated are included in the model's state vector together with physical state variables (see next chapter). In order to compensate for model- and sampling error we inflate all elements of the augmented state vector using spatially and tem-

porally varying adaptive inflation [Anderson, 2007, 2009] and impose a minimum ensemble spread for the parameters. Observations are assimilated four times per day aggregated in 6 hour intervals centered at 0000, 0600, 1200 and 1800 UTC. We use a covariance localization scheme introduced by Gaspari and Cohn [1999] applied in both the horizontal and vertical with a localization half-width of 0.2 rad for the observations and parameters.

The physical part of the state vector comprises the state variables temperature T, horizontal wind speeds U and V, and specific humidity q. Since the tracer q is a positive definite variable, we transform q (in kg kg^{-1}) using

$$\hat{q} = \ln(q) + \alpha \cdot q \tag{1}$$

with $\hat{q}$ being the transformed quantity. The transformation has the advantage that q is still bounded on the lower side via $log(q)$ whereas the linear term dominates for bigger values and keeps the distribution roughly Gaussian rather than log-normal. Two other attempts to transform q did not yield satisfactory results: A simple log-transformation imports the tendency to generate unrealistic large values of q which can lead to a model crash; a transformation based on $\hat{q} = tan(q)$ [Hu et al., 2010] has the advantage of being bounded on two sides but the model did not run stably in test experiments.

2.1. Estimated cloud related parameters

We examine four closure parameters in the cumulus-convection scheme, which we choose in part because the parameter-related processes operate on short timescales. These particular parameters are also known to strongly influence either the model's climate skill or its climate sensitivity [Klocke et al., 2011] and are routinely adjusted during model tuning [Mauritsen et al., 2012]. Table 1 shows the estimated parameters which Tiedtke [1989]

introduced in the mass-flux scheme applied in our model for cumulus convection. The default values for the used model configuration and a range of parameter values as used in different model configurations is given.

All closure parameters are transformed to log-space to avoid possible negative values during the assimilation [Annan et al., 2005a; Tong and Xue, 2008]. Since the Kalman filter assumes Gaussian distributed values of the parameters, they are initialized using a log-normal distribution, which then becomes a Gaussian distribution after the transformation to log-space.

The entrainment-controling parameters ϵ_s and ϵ_p control how much ambient air is mixed into a shallow or deep convective cloud, respectively, and hence influence the cloud's dilution: A high value of entrainment rate imports much surrounding dry air and leads to weaker convection associated with a small vertical extent of the convective plume. Physically, a higher value of ϵ_s increases the entrainment into a shallow convective cloud leading to an increase in cloud water content and generates more stratiform clouds below the inversion. Parameter β gives the fraction of upward moving airmass which overshoots the top of a shallow convective cloud, once it has reached its level of neutral buoyancy and parameter γ describes the conversion rate of liquid water to rain in convective clouds.

Both parameters ϵ_s and β are related to the same process and influence how strongly the boundary layer communicates with the free troposphere across the inversion. Parameter ϵ_s controls the amount of moist air that reaches the inversion and parameter β transports a fraction of the convective airmass across the inversion into the next model level. In all experiments we omit all a priori knowledge and allow the parameters to evolve freely independent of the remaining parameter values.

Unlike conventional state variables in ECHAM6, parameters are usually applied as global constant scalars and need to be treated differently by using the 'spatial updating' method [Aksoy et al., 2006b]. Each ensemble member's scalar parameter is expanded to a horizontally uniform 2D array and spatially updated, yielding a varying posterior field of parameter values. We then feed the spatially weighted global mean back into the model and continue the iterative cycle with the updated prior.

3. Perfect-model experiments

We use perfect-model experiments, sometimes called "observing system simulation experiments" (OSSE) [Atlas, 1997], as a testbed to find a suitable assimilation setup and to explore the individual parameter's sensitivity for a successful parameter estimation. The goal of these idealized experiments is to assess whether the available observations are correlated with the parameters to be estimated, so that also real observations potentially constrain the parameters.

In perfect-model experiments the numerical model is assumed to be perfect in a sense that no model error exists, i.e. a certain set of initial conditions always leads to the same result. Given this assumption, a model run provides a perfect representation of the evolution of the atmosphere. The model is integrated in time and, at constant intervals, the model state is used to generate synthetic observations. As the true state of the model is known, assimilating synthetic observations provides the opportunity to assess the assimilation's performance.

Two sets of perfect-model experiments are performed. The first uses a dense, globally homogeneous distribution of observations (10368 grid cells composed of 9 vertical levels, 24 latitudinal and 48 longitudinal grid-points) allowing for a maximum correlation between

observations and elements of the state vector. The second uses a realistic observation network, identical to the one later used with real observations. Synthetic observations are generated in 6-hour intervals for January 2008 for each observation network. The synthetic observations consist of the horizontal wind speed components U and V, temperature T and specific humidity q. Observational errors for the idealized observation network are constant in space and time with 10 m s^{-1} for both wind components, 10 K for temperature and 2 in log space for specific humidity. For the realistic observation network, we adopt the error specifications of the real observations (for more details see section 5).

For test purposes we conducted experiments with modified observational errors and a different amount of observations, but results were less satisfying in these settings. Reducing the observational error or increasing the amount of available observations leads to a rapid reduction of the distribution spread in the parameter evolution. A collapsed distribution can ultimately lead to filter divergence, i.e. the prior is too confident so that observations are ignored to a large extent making state and parameter estimation impossible.

3.1. Synthetic observations on an idealized observation network

In a most idealized approach (see Figure 1) we estimate each cloud closure parameter in separate perfect-model runs over a period of 30 days, shown in Figure 2(a-d). The solid lines represent the distribution mean and the dashed lines show the distribution width, covering the range of two standard deviations (2σ).

Initial ensemble members for each parameter were drawn from three different specified distributions to assess the impact of the prior ensembles on the robustness of the parameter estimation. The initial mean is chosen such that the different cases clearly exceed the

expert range (Table 1). Each initial distribution spread contains the default parameter value which was set in the run to generate the synthetic observations.

The assimilated observations provide a strong constraint on entrainment-related parameters and a moderate constraint on the buoyancy and microphysical parameters. Closure parameters ϵ_p and ϵ_s in figures 2(a, b) reach a stable state after about 15 days and converge to their default values for all three prior distributions. Closure parameter β in Figure 2(c) also converges, however slower compared to the entrainment rate parameters.

The evolution of conversion rate parameter γ in Figure 2(d) shows a more diverse behavior depending on its initial parameter distribution. The case in which the mean is initialized below the default value converges entirely, whereas the two cases with larger initial mean values exhibit a clear tendency towards the default value but without reaching a stable state after the experiment period of 30 days. One possible reason is, that the first order effect of changes in γ is on precipitation rates. We do not assimilate observations of precipitation, which explains the that this parameter is not well constrained in our experiments.

The faster the distribution mean approaches the default value the smaller the distribution width becomes. Most obvious examples are parameter ϵ_p in Figure 2(b), retaining only a tiny fraction of its initial spread at the end of the experiment (corresponding to the imposed minimum spread), and parameter γ in Figure 2(d) showing a bigger distance to the default value combined with a larger parameter spread.

3.2. Synthetic observations on a realistic observation network

We repeat the previous experiments but using both the observation network and error specifications used in the next section. Using a realistic observation network produces

qualitatively similar results as experiments with an idealized observation network (compare Figure 2 and 3). Parameters ϵ_s, ϵ_p and β converge to the default value for all chosen initial parameter distributions. However, the rates of convergence differ slightly among the parameters, especially two cases of ϵ_p converge at a slower rate to the default value. Parameter γ also exhibits a slower convergence rate in one case, without reaching the default value during the experiment, and estimation is not successful if initialized with a distribution far above the default value. Even though a perfect model is utilized, experiments with parameter γ show that an entirely successful estimation of closure parameters depends on properties of the observation network such as the spatial distribution of observations and the specified observational error.

Overall, the results suggest that the observations assimilated here can constrain the estimated parameters and that parameter settings for climate simulations can be recovered by optimizing cloud related processes for short forecasts. Both entrainment parameters and the mass flux parameter can be estimated more robustly than conversion-rate parameter γ. This is not surprising as the other parameters have a more direct control on the thermodynamic state of the atmosphere. In fact, ϵ_p has the largest control on the skill of the simulated climate [Klocke et al., 2011], while also being effectively constrained by short weather forecasts.

4. Adding complexity

We take small steps towards estimating parameters with real observations to be able to understand the final results which can not be verified. We use two setups (Figure 1): In a first approach we add complexity by estimating all four parameters simultaneously, but still operate in the perfect-model setup using the realistic observation network

(see section 3.2). The second experiment type imitates model deficiency in the forecast model. We compare these two experimental setups with the parameter evolution in the perfect-world setting from the previous section 3.2, in which parameters are estimated individually. The three experiment types assimilate identical synthetic observations on a realistic observation network and are plotted in Figure 4.

4.1. Estimating parameters simultaneously

All parameters converge to their default values in the perfect model when estimated one at a time as well as when estimated simultaneously (Figure 4, blue and grey lines). The rate of parameter convergence is slightly affected when estimated together with other parameters. No difference between the two model setups is observed for parameter ϵ_s in Figure 4(a), whereas ϵ_p and γ reach a stable state even faster when all four parameters are estimated at the same time. Parameter β converges slightly slower in the multiple parameter estimation than when estimated alone. In contrast to the experience of Aksoy et al. [2006a] and Tong and Xue [2008] the success of the parameter estimation does not suffer, when increasing the number of estimated parameters. However, we note that estimating parameters simultaneously increases the potential of unstable model states which lead to model crashes. This behaviour is strongly dependent on the choice of initial parameter distributions.

4.2. Imperfect models

In a second experiment type we introduce model imperfectness by setting the gas constant R to 15 J mol^{-1} K^{-1} (rather than 8.31 J mol^{-1} K^{-1}) and gravity g to 6.0 m s^{-2} (rather than 9.81 m s^{-2}). Both changes have relatively little impact on the models perfor-

mance on short and long timescales. In this experiment all four parameters are estimated simultaneously. Compared to the previous experiments with a perfect model, estimation performance is degraded when employing an imperfect model. All parameters still show a tendency of convergence but distant from the truth. This is to be expected as effects of the altered gravity and gas constant on the short forecasts are compensated for by altering the parameters, to still achieve the best fit to the synthetic observations from the default model configuration.

In the case of multiple parameter estimation in an imperfect-model setup, we can explain the joint behavior of parameters ϵ_s and β by taking their physical relationships into account. Both parameters control the communication between the boundary layer and the free troposphere across the inversion and are related to the same process. Since we assimilate synthetic observations, the amount of mass flux across the inversion is given by the combination of the default parameter values of ϵ_s and β. The lack of convergence to the 'true' value when parameters are estimated simultaneously in the imperfect-model setup can be explained by compensating effects. Similar compensating relations between the remaining parameters and, secondly, the introduction of an imperfect model lead to parameter convergence distant from the truth.

These findings remind us to be cautious when confronting the incomplete climate model with the real world and interpreting the results. We can not expect a single universally applicable value for the parameters, but the best fit of a certain model configuration to the observations at hand. To achieve the best fit to observations we ask the parameterizations to compensate for missing processes, or structural errors in the model.

4.3. Which observations inform the parameters?

Since parameters β and γ show a weak estimation robustness in perfect-model experiments, we analyze the parameter's sensitivity to different observation quantities using the idealized observation network. Examining the assimilation performance we seek the degree to which each observation quantity constrains the individual parameters.

For parameters β and γ we perform experiments corresponding to the four available observation quantities U, V, T and q, which are exclusively assimilated in separate runs. In order to compare results, the initial parameter distributions are identical in each parameter case.

Parameter β shows sensitivity to all observation quantities but assimilation of specific humidity leads to the best and temperature to the least optimal results (Figure 5(a)). Furthermore, we find that assimilating only specific humidity leads to a better result than using all observation quantities, by comparison to Figure 2(c).

Figure 5(b) shows that parameter γ is mostly constrained by U and q. The sensitivity of γ to V appears to be low, leading to a slightly increasing evolution of the distribution mean and a growing distribution spread in this case. Using T as the only observation leads to a strongly diverging mean and increasing spread, implying that γ and T are entirely uncorrelated. We suggest that an uncorrelated estimation of a single bounded quantity results in a growing upper tail on the unbounded side of the distribution. We therefore observe a statistical artifact rather than a substantial and envisaged increase of the distribution mean and spread.

5. Parameter estimation using real observations

The previous experiments demonstrate that the observations available to us can, at least in principle, constrain the parameters we seek to estimate to some degree. Here we

turn to the practical task of estimation using real observations, against which our model is imperfect. We assimilate observations of T, U, V and q measured by radiosondes, aircrafts and satellites (only U and V) in January 2008. Observation errors and locations come from the metadata describing the observations used in the NCEP Reanalysis [Kistler et al., 2001]. We estimate each parameter independently, using four initial distributions each (Figure 6).

Experiments of parameter ϵ_s prefer a range of values centered around $1 \cdot 10^{-3}$ m^{-1} without converging to a distinct single value. Experiments with large eddy simulations [Siebesma and Cuijpers, 1995] suggest higher values for ϵ_s with $1.5-2 \cdot 10^{-3}$ m^{-1}, indicating a better representation of cloud phenomena on short timescales. In contrast, Tiedtke [1989] suggest a lower value of $3 \cdot 10^{-4}$ m^{-1} which is often employed in climate models. Järvinen et al. [2010] also estimate ϵ_s in ECHAM5 with an adaptive MCMC technique, evaluating a cost function based on radiative fluxes at TOA. Even though they use a substantially different method and a cost function based on other quantities they estimate ϵ_s to $1.5 \cdot 10^{-3}$ m^{-1}, close to the upper end of our final distribution of parameter values.

The most strongly constrained parameter is ϵ_p, which converges to a value about $8 \cdot 10^{-5}$ m^{-1}, c.f. the default value of $1 \cdot 10^{-4}$ m^{-1}, regardless of the parameter's initial distribution. Compared to the other parameters, experiments with ϵ_p yield the most robust evolution. This confirms previous results that ϵ_p is strongly correlated with the available observations. However, this parameter is not entirely stable at the end of the experiment period and seems to continue to drift towards smaller values.

By definition of the parameterization, β is a quantity that should be bounded between 0 m^{-1} and 1 m^{-1}. Since parameters are transformed to log space, the assimilation can

practically produce values larger than 1 m^{-1}. Given this a priori information combined with the fact that the estimate exceeds the upper limit at 1 m^{-1} draws a first conclusion that the assimilation of real observations fails for parameter β.

As is evident from the perfect-model results, parameter γ is poorly constrained by in-situ measurements. Different initializations of parameter γ clearly diverge but all estimates indicate parameter values at the upper end or above the expert range. We conclude that correlations between available real observations and γ do not suffice to constrain or estimate this parameter robustly.

In view of the parameter sensitivity to different observation quantities (Figure 5) we repeat experiments for β and γ, neglecting uncorrelated observation types like T in some experiments and exclusively assimilating correlated quantities like q in other experiments [Kang et al., 2011]. Results with selected observation quantities (not shown) confirm the nature of the previous results in which all observation quantities are assimilated.

6. Do better parameters lead to a more faithful model?

6.1. Tests at short timescales

Though data assimilation assures that the parameter values estimated are optimal, i.e. are most consistent with the observations subject to error estimates of both observations and parameters, there is no guarantee that a model with updated parameters will have lower errors in state variables. Since there is no 'true' state available to assess the estimation performance as in perfect-model experiments, we are left with calculating the root-mean square error (RMSE) between each assimilated observation and the models 6-hour forecast interpolated to observation space. Assimilating real observations yields new and reasonable values for parameters ϵ_s and β; we neglect the remaining parameters

because the estimate produces a value nearly identical to the default value (ϵ_p) or diverges (γ).

Figure 7 shows the error distribution, averaged over space and time but resolved in the vertical, for the northern hemisphere, the tropics, the southern hemisphere, and a global average. We show three observation types and note that error distributions for other observation sources are similar. We omit showing RMSE for q since only few observations are available which provide an unrepresentative error distribution due to possible sampling error. In each subplot we compare three experiments with different parameter estimation setups: State estimation with constant default parameter values, simultaneous state and parameter estimation for ϵ_s and β.

Comparing the default parameter setting with the simulation results applying the estimated parameter setting, both parameters lead to a decreased RMSE for the zonal and meridional wind observation types for different geographical regions. In the southern hemisphere however, the assimilation of aircraft meridional wind shows both an error reduction and an error increase, depending on the vertical level. Looking at the RMSE of radiosonde temperature, the updated parameters have little effect in the northern hemisphere and both positive and negative effect in the tropics and the southern hemisphere.

The overall impact of updated parameters on global average RMSE of 6-hour forecasts shows a neutral (ϵ_s) and a marginal positive (β) effect for radiosonde temperature. Both wind components show a clear error reduction with estimated parameters.

6.2. Tests at climatological timescales

Parameter estimation reduces the short term model forecast error by optimizing fast processes. This brings up the question whether results from parameter estimation on

short timescales are transferable to climatological timescales: Does the representation of the mean climate state also improve with a presumably better representation of physical processes?

We conduct climate simulations with parameter values obtained in section 5 and assess the model's performance. The performance index I^2 [Reichler and Kim, 2008] constitutes a quantification of the agreement between model and observations in an integrated quantity. The index consists of the aggregated error in simulating the observed climatological means of relevant climate observables. In contrast to the original composition of observables in Reichler and Kim [2008] we use a reduced set of observations, described in [Stevens et al., 2012].

Figure 8 shows I^2 for different regions and parameter settings using $1 \cdot 10^{-3}\ m^{-1}$ for ϵ_s and $1\ m^{-1}$ for β, keeping ϵ_p and γ at their default values. The runs with the updated parameter values do not improve the model's performance on climatological scales compared to the default setting. This can have several reasons: a) The filter tries to compensate for systematic model errors by suggesting unrealistic parameter values, b) parameters are optimized to give a good 6h forecast, omitting interacting processes and feedbacks which can occur on longer timescales, c) the chosen model setup with fixed SST and SIC strongly constrains the model's variability and dominates most components of I^2, d) looking at the geographical composition of I^2 worst performance is achieved in the southern hemisphere where the lack of sufficient observations may hinder a correct estimation of global scalar parameters (however, at short timescales, we found a particular improvement in the southern hemisphere), e) model errors in the skill metrics applied for NWP and climatological timescales, which comprise different variables, do not correlate.

7. Data assimilation and the tuning of climate models

7.1. Results and utility

The perfect-model experiments show that EAKF successfully estimates four cloud closure parameters in a comprehensive atmospheric GCM (ECHAM6). Practically this means that covariance relations exists between the available synthetic observations (U,V,T,q) and the closure parameters. However, not only the amount and error of synthetic observation but also filter settings like inflation need to be tuned to achieve a successful parameter estimation. This exercise has been exploratory and we expect that further development of the assimilation system may yield better, if not fundamentally different, results.

Different parameters are estimated with varying degrees of success. Entrainment parameters ϵ_s and ϵ_p are sufficiently correlated with observations in all experiments and are robustly estimated, while estimation of γ succeeds when assuming a perfect model, but reveals difficulties when increasing the imperfectness of the model and fails in an assimilation framework with real observations. The success with which a parameter can be estimated is consistent across the hierarchy of these experiments: Entrainment-rate parameter ϵ_p shows the highest rate of convergence in perfect-model experiments and estimation succeeds with real observations, while conversion-rate parameter γ shows the lowest rate of convergence and estimation fails with real observations. Using additional observation types like precipitation might ameliorate the estimation performance of certain parameters. However, critical issues remain such as acquiring trustworthy observations and assimilating bounded observations which exhibit a non-Gaussian distribution.

We demonstrate in experiments with synthetic observations that different observation quantities constrain the parameters to a different extent (Figure 5). Analogously, we conduct parameter estimation experiments with real observations in which the different observation quantities are assimilated individually (not shown). Independent of the observation quantity, both experiments show a similar behavior in the parameter evolution. We therefore conclude that the limiting factor for a successful parameter estimation with real observations is not the quality of the observations, but rather generic model deficiencies, provided that assimilated observations constrain the estimated parameters.

7.2. Technical issues

Estimation performance of parameter β demonstrates that the assimilation of bounded quantities within an EnKF assimilation framework is not an entirely solved issue yet. The transformation with a function based on the logarithm alleviates the problem for single-bounded quantities but simultaneously introduces non-Gaussian distributions, an undesired distribution property for EnKF. Nevertheless, results from logarithmic-based transformations of q, ϵ_{p} and ϵ_{s} prove that the logarithm is a feasible option for single-bounded quantities. The overshooting of the upper limit of the double-bounded quantity β in Figure 6(c) shows that this aspect of the EnKF data assimilation requires further work [e.g. Anderson, 2010]. In a first attempt to find a formulation for double-bounded quantities, we apply transformations based on $\hat{q} = tan(q)$ but experience unstable model states in the posterior.

As a technical aspect of this study we see that expanding the scalar parameter to a 2D array, updating the parameter spatially and finally averaging is a feasible approach. A spatially varying parameter field in the posterior yields potential information about

the preferred geographical distribution of parameter values. These maps can help to construct and evaluate the feasibility of a spatially varying cloud parameterization in weather and climate models. The parameter map also offers the possibility to weight or restrict parameters to a certain geographical region, e.g. ϵ_s should be particularly important in trade wind regimes. Furthermore, it is technically easy to expand the scalar parameter to a more comprehensive 3D field for the assimilation which might produce even better estimation results.

Estimating parameters in short time-scale parameterizations is computationally more efficient than estimating them using quasi-climatological integrations. For example, Järvinen et al. [2010] estimate three of the four parameters with a Markov Chain Monte Carlo (MCMC) approach which requires 4500 years of model simulations in their case. Our method provides a nearly identical result for parameter ϵ_s with 90 model members each integrated 30 days. If the same model resolution is utilized, the EAKF method therefore requires only 0.17% of the computational time of MCMC to integrate the model in time.

7.3. Applicability

It remains an open question why better estimates on short timescales lead to a deteriorated performance on climatological timescales. To understand this discrepancy is of vital importance for the future work on the estimation of climate model parameters in a NWP context. We present possible explanations in section 6.2 but further investigation is necessary in order to isolate the dominating reasons.

Acknowledgments. We thank the Max Planck Society, the International Max Planck Research School for Earth System Modeling, the National Science Foundation's Center

for Multi-scale Modeling of Atmospheric Phenomena, the German Research Foundations Emmy Noether grant and the European Union, Seventh Framework Program EUCLIPSE for supporting this work. Simulations were carried out on the supercomputing facilities of the German Climate Computation Centre (DKRZ) in Hamburg. Lorenzo Tomassini and Bjorn Stevens provided valuable feedback on earlier drafts of this work and Thomas Reichler provided the code to calculate I^2 to evaluate the climate simulations.

References

Aksoy, A., F. Zhang, and J.W. Nielsen-Gammon, Ensemble-Based Simultaneous State and Parameter Estimation in a Two-Dimensional Sea-Breeze Model, Mon. Wea. Rev., 134, 2951–2970 (2006).

Aksoy, A., F. Zhang, J.W. Nielsen-Gammon, and C.C. Epifanio, Ensemble-based simultaneous state and parameter estimation with MM5, Geophys. Res. Lett., 33, L12801 (2006).

Allen, M. R., Do-it-yourself climate prediction. Nature, 401, 627 (1999).

Anderson, J. L., An Ensemble Adjustment Kalman Filter for Data Assimilation, Mon. Wea. Rev., 129, 2884–2903 (2001).

Anderson, J. L., An adaptive covariance inflation error correction algorithm for ensemble filters, Tellus A, 59, 210–224 (2007).

Anderson, J. L., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, The Data Assimilation Research Testbed: A Community Facility, Bull. Am. Meteorol. Soc., 90, 1283–1296 (2009).

Anderson, J. L., Spatially and temporally varying adaptive covariance inflation for ensemble filters, Tellus A, 61, 72–83 (2009).

Anderson, J. L., A Non-Gaussian Ensemble Filter Update for Data Assimilation, Mon. Wea. Rev., 138, 4186–4198 (2010).

Annan, J.D., and J.C. Hargreaves, Efficient parameter estimation for a highly chaotic system, Tellus A, 56, 520–526 (2004).

Annan, J. D., D. J. Lunt, J. C. Hargreaves, and P. J. Valdes, Nonlinear Processes in Geophysics Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter, Nonlin. Processes Geophys., 363–371 (2005).

Annan, J.D., J.C. Hargreaves, N.R. Edwards, and R. Marsh, Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter, Ocean Modelling 8, 135–154 (2005).

Atlas R., Observing System Simulation Experiments: methodology, examples and limitations, Proceedings of CGC/WMO Workshop, 7–9 (1997).

Bony, S., et al. How well do we understand and evaluate climate change feedback processes? J. Climate, 19, 3445–3482 (2006).

Derber, J. C., A variational continuous assimilation technique, Mon. Wea. Rev., 117, 2437–2446, (1989).

Evensen, G., The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dyn., 53, 343–367 (2003).

Gaspari, G., and S. E. Cohn, Construction of correlation functions in two and three dimensions, Quart. J. Royal Meteorol. Soc., 125, 723–757 (1999).

497 Stevens, B., et al., Climate variability and climate change in MPI-ESM simulations. *In*
498 *preparation* (2012).

499 Hu, X.M., F. Zhang, and J.W. Nielsen-Gammon, Ensemble-based simultaneous state
500 and parameter estimation for treatment of mesoscale model error: A real-data study,
501 Geophys. Res. Lett., 37, L08802 (2010).

502 Jackson, C. S., M. K. Sen, G. Deng, and K. P. Bowman, Error Reduction and Convergence
503 in Climate Prediction, J. Climate, 21, 6698–6709 (2008).

504 Järvinen, H., P. Räisänen, M. Laine, J. Tamminen, A. Ilin, E. Oja, A. Solonen, and
505 H. Haario, Estimation of ECHAM5 climate model closure parameters with adaptive
506 MCMC, Atmos. Chem. Phys., 10, 9993–10002 (2010).

507 Kang, J.-S., E. Kalnay, J. Liu, I. Fung, T. Miyoshi, and K. Ide, 'Variable localization'
508 in an ensemble Kalman filter: Application to the carbon cycle data assimilation, J.
509 Geophys. Res., 116, D09110 (2011).

510 Kistler, R., W. Collins, S. Saha, G. White, J. Woollen, E. Kalnay, M. Chelliah, W.
511 Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne, and M. Fiorino,
512 The NCEP-NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation,
513 Bull. Am. Meteorol. Soc., 82, 247–267 (2001).

514 Klocke, D., R. Pincus, and J. Quaas, On Constraining Estimates of Climate Sensitivity
515 with Present-Day Observations through Model Weighting, J. Climate, 24, 6092–6099
516 (2011).

517 Knutti, R., T. F. Stocker, F. Joos, and G. -K. Plattner, Constraints on radiative forcing
518 and future climate change from observations and climate model ensembles, Nature, 416,
519 719–723 (2002).

520 Mauritsen, T., B. Stevens, E. Roeckner, T. Crueger, M. Esch, M. Giorgetta, H. Haak, J.
521 Jungclaus, D. Klocke, D. Matei, U. Mikolajewicz, D. Notz, R. Pincus, H. Schmidt, and
522 L. Tomassini, Tuning the climate of a global model, Submitted to JAMES (2012).
523 Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins,
524 and D. A. Stainforth, Quantification of modelling uncertainties in a large ensemble of
525 climate change simulations, Nature, 430, 768–772 (2004).
526 Norris, P. M., and A. M. da Silva, Assimilation of Satellite Cloud Data into the GMAO
527 Finite-Volume Data Assimilation System Using a Parameter Estimation Method. Part
528 I: Motivation and Algorithm Description, J. Atmos. Sci., 64, 3880–3895 (2007).
529 Randall, D. A., and B. A. Wielicki, Measurements, models and hypotheses in the atmo-
530 spheric sciences, Bull. Am. Meteorol. Soc. 82, 283–294 (1997).
531 Reichler, T., and J. Kim, How well do climate models simulate today's climate? Bull.
532 Am. Meteorol. Soc., 89, 303–311 (2008).
533 Rodwell, M. J., and T. N. Palmer, Using numerical weather prediction to assess climate
534 models, Quart. J. Royal Meteorol. Soc., 133, 129–146 (2007).
535 Siebesma, A. P., and J. W. M. Cuijpers, Evaluation of parametric assumptions for shallow
536 cumulus convection, J. Atmos. Sci., 52, 650–666, (1995).
537 Solomon, S., et al., Climate Change 2007: The Physical Science Basis. Contribution of
538 Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on
539 Climate Change, Cambridge University Press, Cambridge, United Kingdom and New
540 York, USA (2007).
541 Soden, B. J., and I.M. Held, An assessment of climate feedbacks in coupled ocean-
542 atmosphere models, J. Clim., 19, 3354–3360 (2006).

Tiedtke, M. A., Comprehensive mass flux scheme for cumulus parameterization in large-scale models, Mon. Wea. Rev., 117, 1779–1800 (1989).

Tong, M., and M. Xue, Simultaneous Estimation of Microphysical Parameters and Atmospheric State with Simulated Radar Data and Ensemble Square Root Kalman Filter. Part II: Parameter Estimation Experiments, Mon. Wea. Rev., 136, 1649–1668 (2008).

Idealized observations ↓ *Real observations*

Observation type	Observation network	Parameter estimation	Model
Synthetic	Idealized	Single	Perfect
Synthetic	Real	Single	Perfect
Synthetic	Real	Multiple	Perfect
Synthetic	Real	Single	Semiperfect
Real	Real	Single	Imperfect

Perfect model setup ↓ *Imperfect model setup*

Figure 1. Hierarchy of experimental setups with decreasing degrees of idealization: Bridging from experiments with synthetic observations on an idealized observation network, assuming a perfect-forecast model, to real observations and a consequently imperfect model.

Parameter	Acronym	Range	Default value	Unit
Entrainment rate for shallow convection	ϵ_s	$3 \cdot 10^{-4} - 1 \cdot 10^{-3}$	$3 \cdot 10^{-4}$	m^{-1}
Entrainment rate for penetrative convection	ϵ_p	$3 \cdot 10^{-5} - 5 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	m^{-1}
Cloud mass flux above level of non-buoyancy	β	0.1-0.3	0.27	m^{-1}
Conversion rate from cloud water to rain	γ	$1 \cdot 10^{-4} - 5 \cdot 10^{-3}$	$4 \cdot 10^{-4}$	s^{-1}

Table 1. List of closure parameters with the corresponding default values for ECHAM6 at T31L19. The range of parameter values is chosen by expert elicitation and has been used in parameter perturbation experiments by Klocke et al. [2011] with ECHAM5.

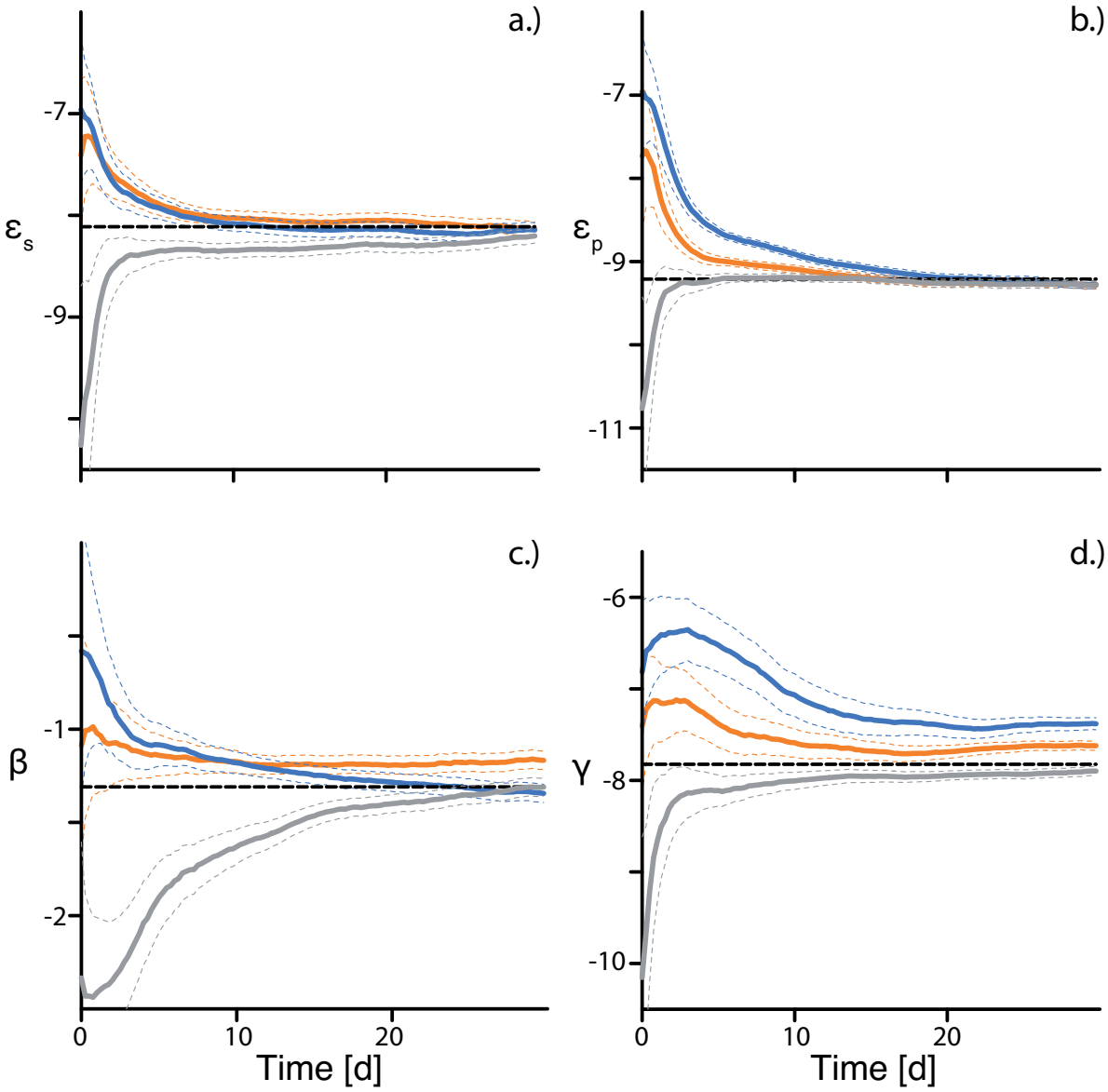

Figure 2. Evolution of closure parameters when estimated individually in a 'perfect model' experiment assimilating observations on an idealized network. Three different initial distributions (colors) are shown by the distribution mean (solid lines) and spread (dashed lines). The parameter values are shown in log space and the dashed black line displays the default parameter value which was set in the run to generate the synthetic observations.

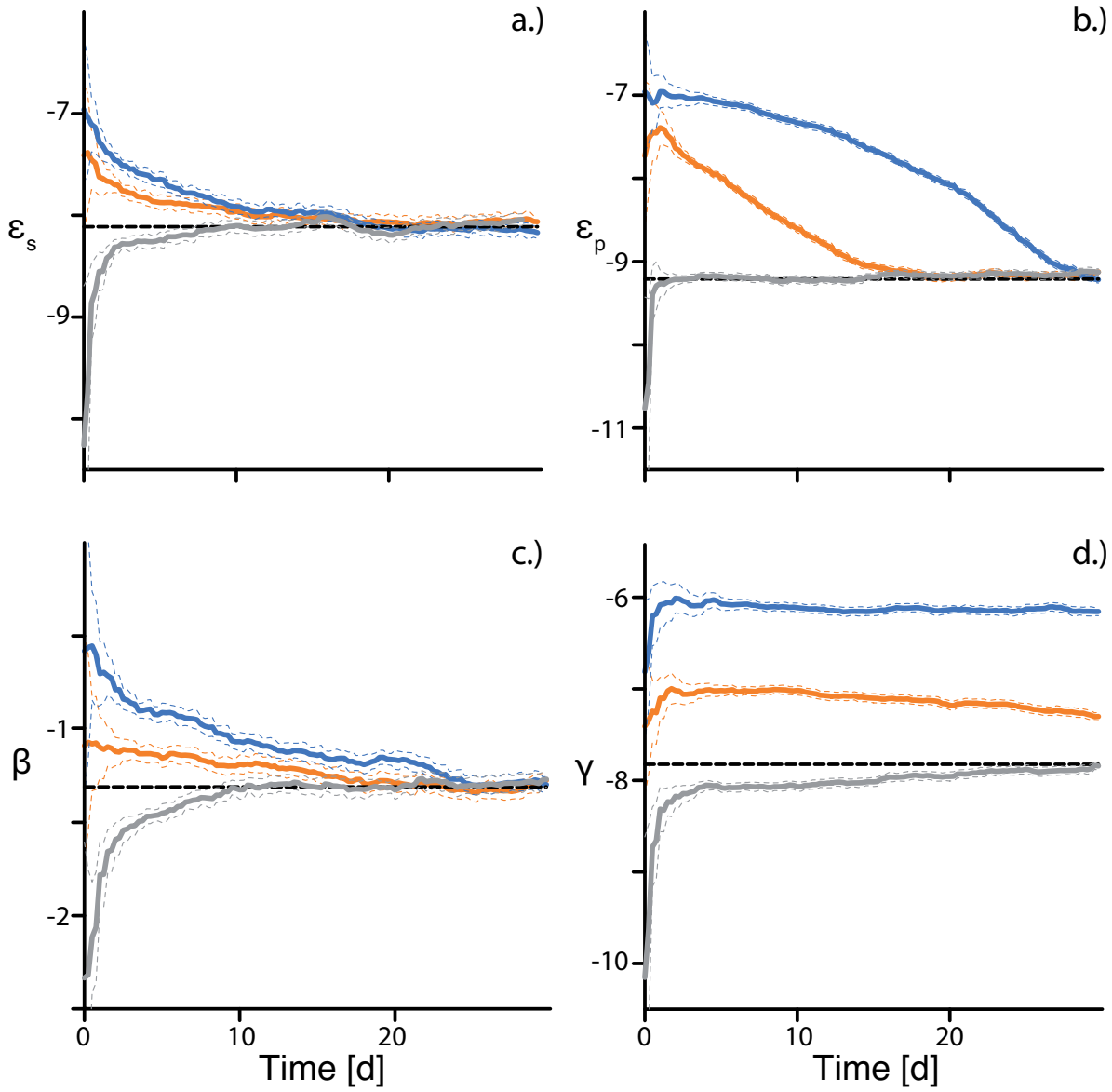

Figure 3. Same as figure 2 but assimilating observations on the realistic observation network.

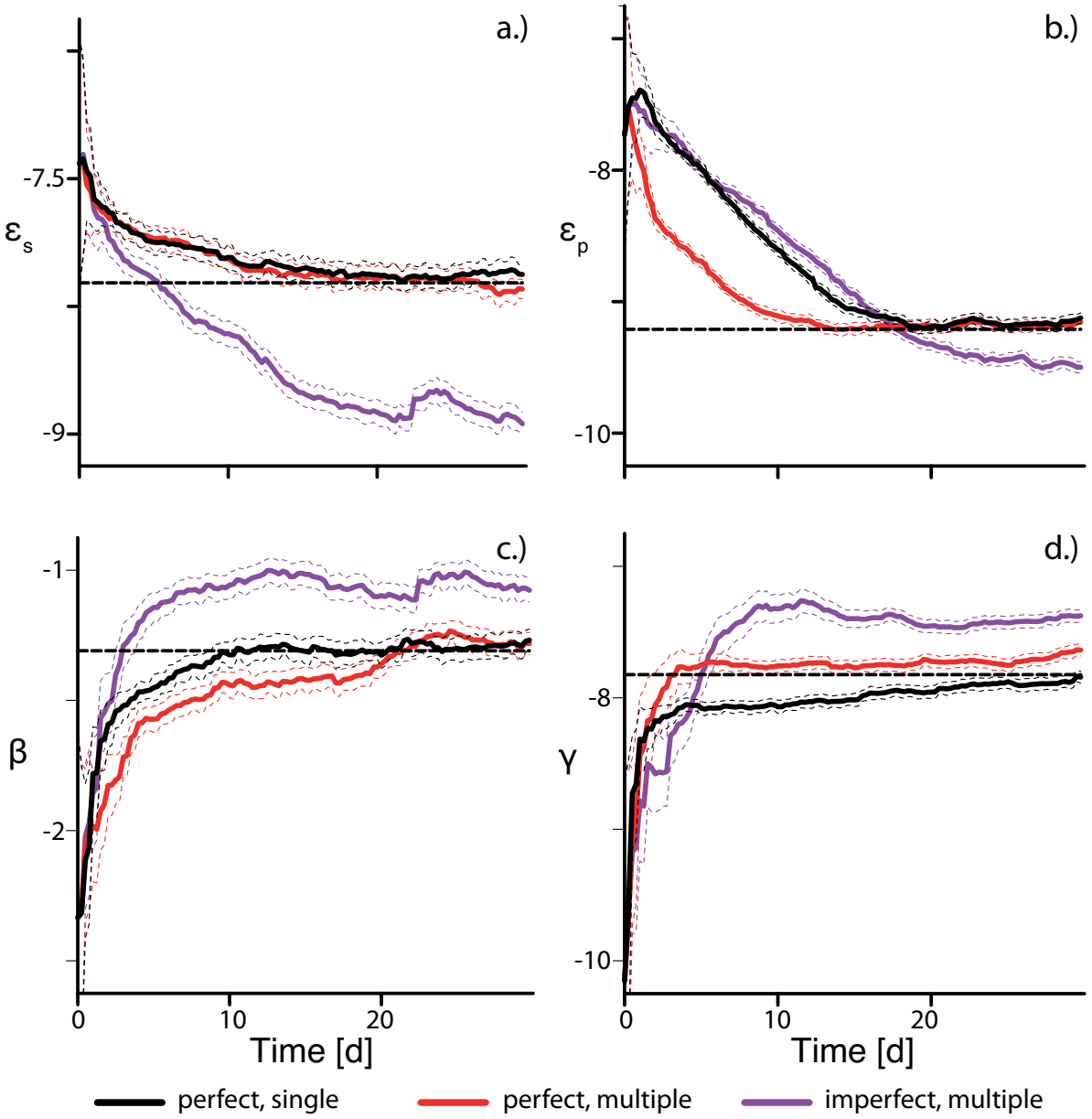

Figure 4. Evolution of closure parameters with different model settings showing the distribution mean (solid lines) and spread (dashed lines). The parameter values are shown in log space and the dashed black line displays the default parameter value which was set in the run to generate the synthetic observations on a homogeneous observing network. Three different model setups are displayed: perfect model results with single parameter estimation (black), four parameters estimated simultaneously (red) and four parameters estimated simultaneously with perturbed gas constant and gravity (purple).

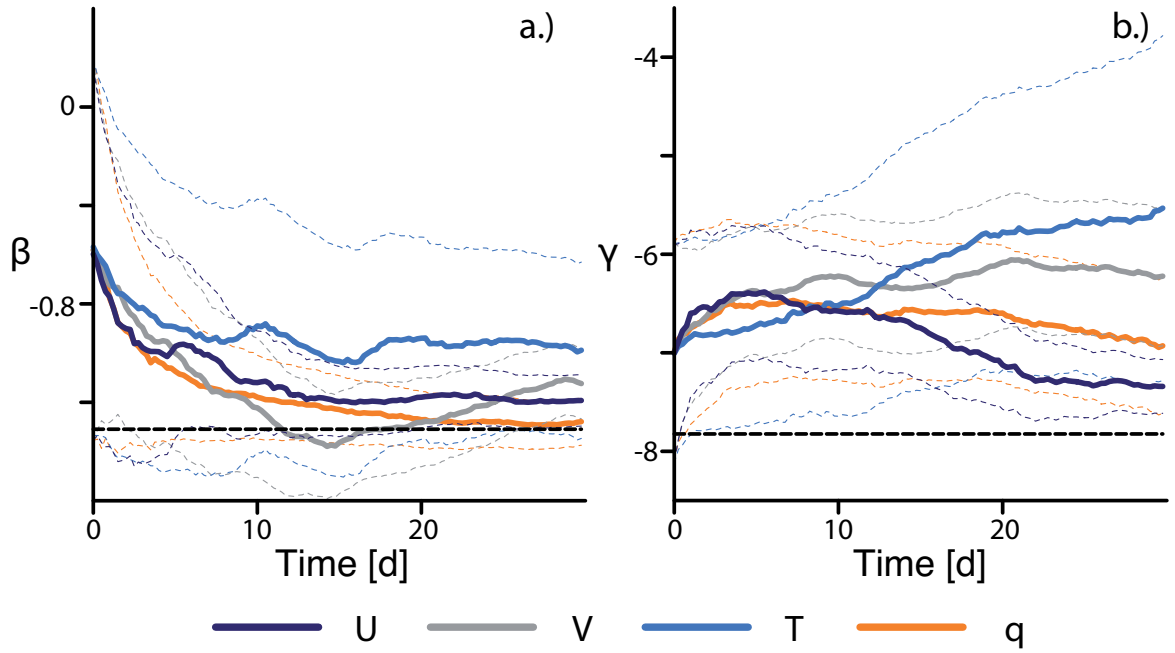

Figure 5. Sensitivity of closure parameters to different observation quantities, showing the distribution mean (solid lines) and spread (dashed lines). Parameters are shown in log space and the dashed black line displays the default parameter value which was set in the run to generate the synthetic observations on a homogeneous observing network. Experiments assimilating only U, V, T, and q are shown in purple, grey, blue, and orange, respectively.

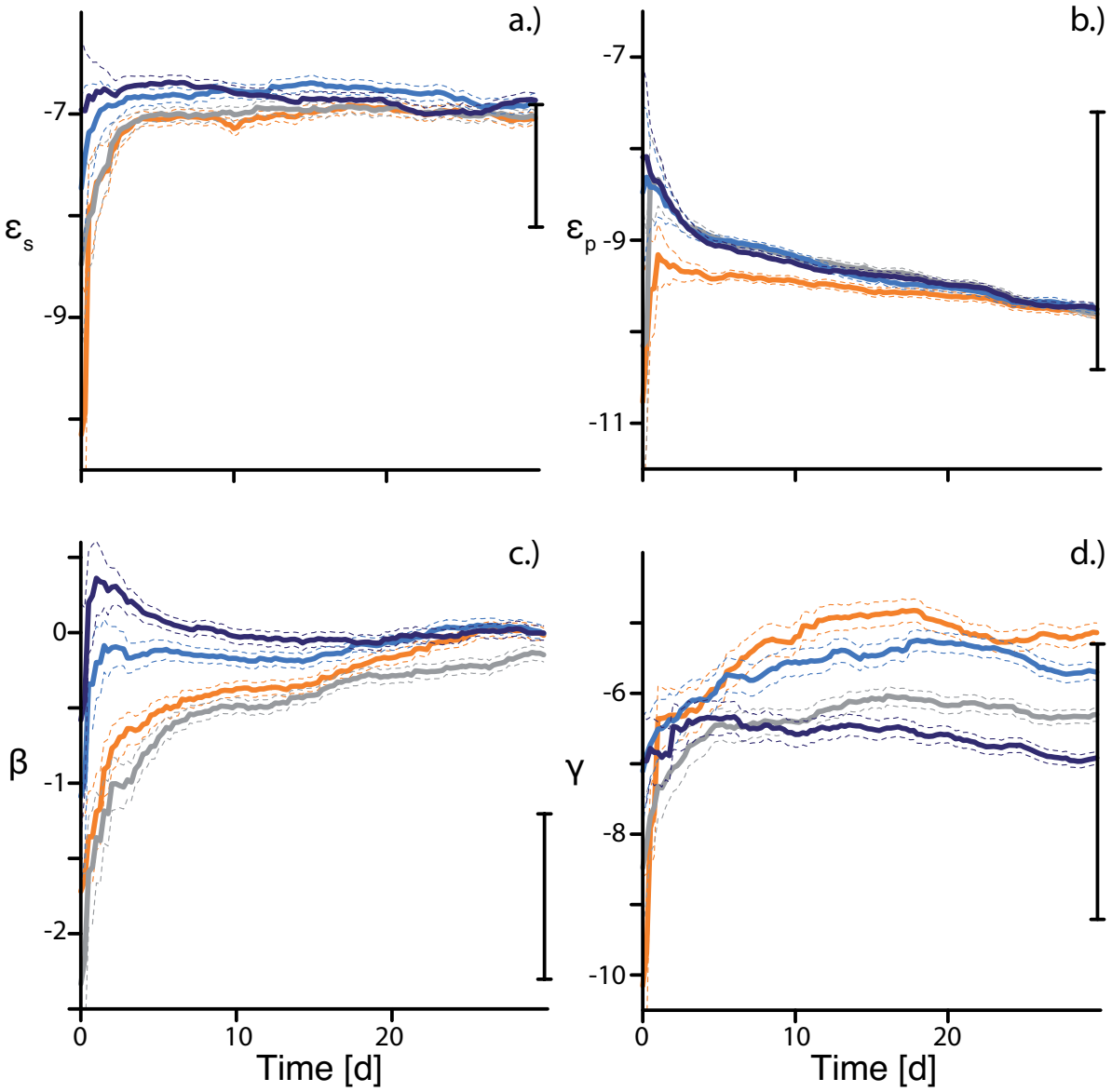

Figure 6. Evolution of closure parameters with assimilation of real observations showing the distribution mean (solid lines) and spread (dashed lines) for different initial parameter distributions (colors). The parameter values are shown in log space and the vertical black range displays possible parameter values chosen by expert elicitation.

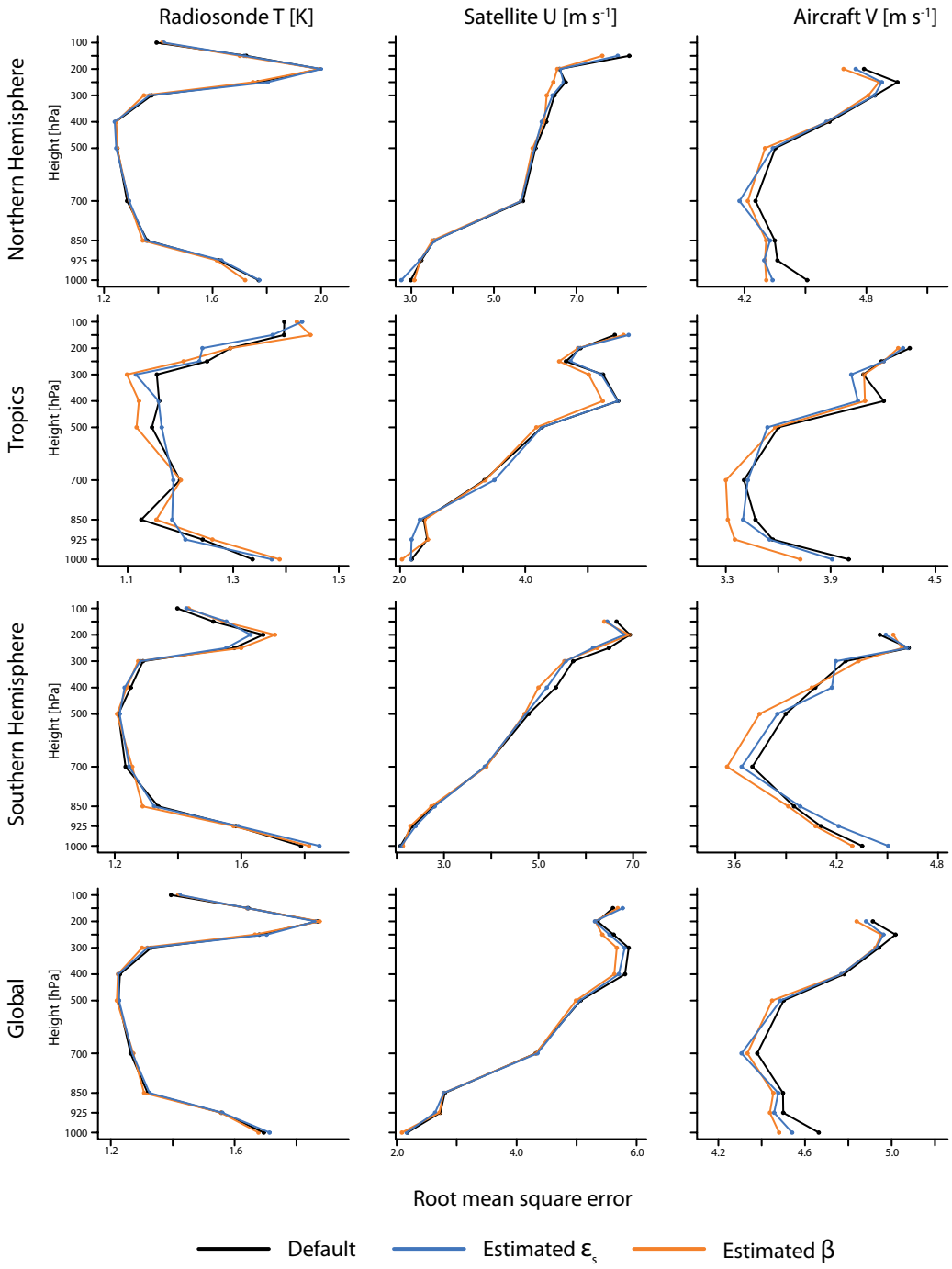

Figure 7. Error estimate for experiments with real observations for different geographical regions (rows) and different observation types (columns). Each observation type is regarded as 'truth' to calculate the root-mean-square error (RMSE). Each subplot contains a horizontal and time averaged vertical profile with default parameter setting (black), estimated ϵ_s (blue) and estimated β (orange). In the course of the experiment parameter ϵ_s converges to $1 \cdot 10^{-4}\ m^{-1}$ and parameter β to $1\ m^{-1}$.

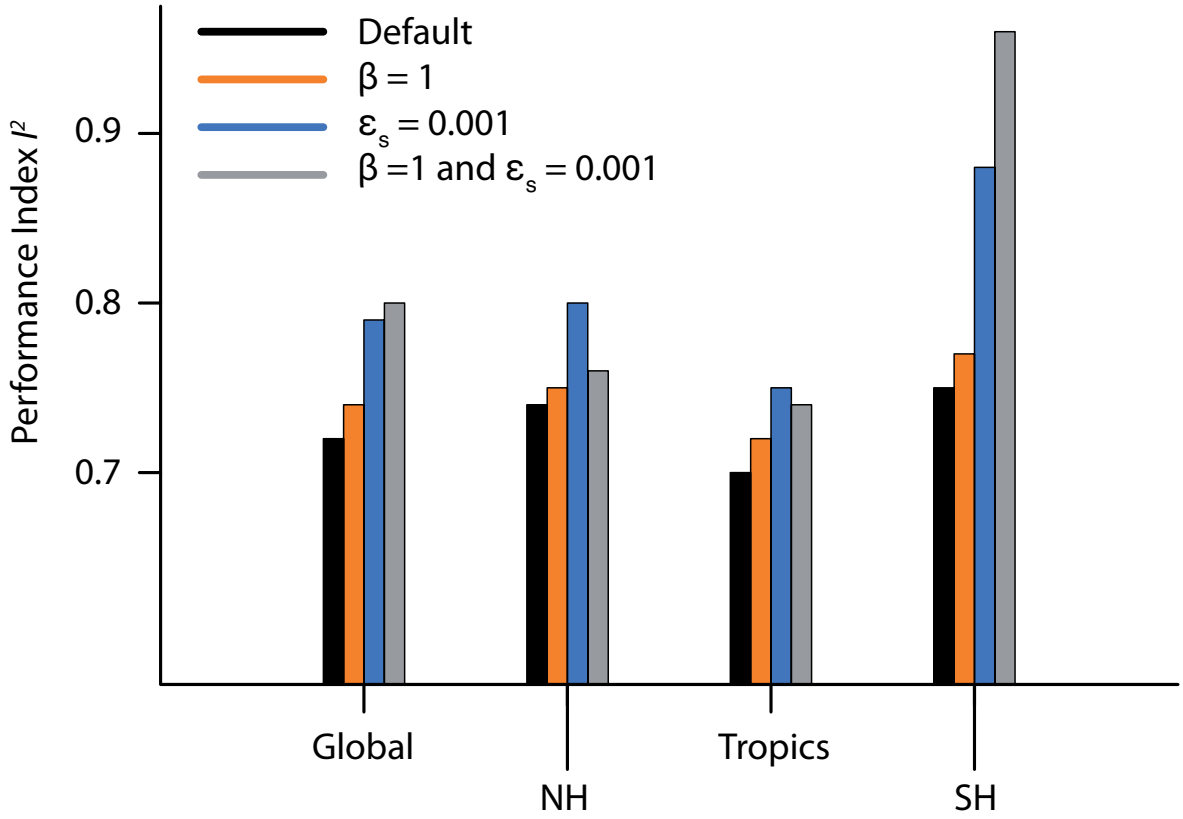

Figure 8. Aggregated error of simulated climate runs, using four different parameter settings. A run with default parameter setting is compared to runs in which ϵ_s and β are changed individually and simultaneously. Four geographical regions are shown: the global average, the northern (NH) and southern hemisphere (SH) and the tropics. A low I^2 value corresponds to a good performance. In contrast to the original composition of observables, a reduced set of observations is used, described in [Stevens et al., 2012].